

GenCore version 5.1.6
Copyright (c) 1993 - 2005 Compugen Ltd.

OM protein - nucleic search, using frame_plus_p2n model
Run on: December 25, 2005, 02:03:13 ; Search time 1154 Seconds
(without alignments)
3618.749 Million cell updates/sec

Title: US-10-763-712A-26
Perfect score: 2604
Sequence: 1 M A L G L A E L R A G A V A C A R R H R A H E L L H T Y P G A R A D A 505

Scoring table: BLOSUM62
Xgapop 10.0 , Xgapext 0.5
Ygapop 10.0 , Ygapext 0.5
Fgapop 6.0 , Fgapext 7.0
Delop 6.0 , Delext 7.0

Searched: 9793542 seqs, 4134689005 residues
Total number of hits satisfying chosen parameters: 19587084

Minimum DB seq length: 0
Maximum DB seq length: 2000000000
Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Command line parameters:
-MODEL=frame+ p2n.model -DEV=xlh
-Q=/cgn2_1/USPTO.spool/US10763712/runat_21122005_164206_10308/app_query.fasta_1.647
-DB=Published Applications NA.Main -OPMT=fastap -SUFFIX=p2n.rnpbm
-MINMATCH=0.1 -LOOPCL=0 -LOOPT=0 -UNITS=bits -START=1 -END=1
-MATRIX=BLOSUM62 -TRANS=human40.cd1 -LIST=45 -LOCAL=LOCAL -OUTFMT=pt0 -NORM=ext
-THR_MAX=100 -THR_MIN=0 -ALIGN=15 -MODE=LOCAL -OUTFMT=pt0 -NORM=ext
-HEAPSIZE=500 -MINLEN=0 -MAXLEN=2000000000
-USER=US10763712 @CEN 1.1 1026 @runat_21122005_164206_10308 -NCPU=6 -ICPU=3
-NO MMAP -LARGEQUERY -NEG SCORES=0 -WAIT -DSPLOCK=100 -LONGLOG
-FGAPEXT=7 -YGAPOP=10 -WARN_TIMECUT=30 -THREDS=1 -XGAPOP=10 -XGAEXT=0.5 -FGAPOP=6
-XGAEXT=7 -YGAPOP=10 -XGAEXT=0.5 -DELOP=6 -DELEXT=7

Database : Published Applications NA Main:
1: /cgn2_6/ptodata/1/pubpna/US07_PUBCOMB.seq:
2: /cgn2_6/ptodata/1/pubpna/US08_PUBCOMB.seq:
3: /cgn2_6/ptodata/1/pubpna/US09A_PUBCOMB.seq:
4: /cgn2_6/ptodata/1/pubpna/US09B_PUBCOMB.seq:
5: /cgn2_6/ptodata/1/pubpna/US10A_PUBCOMB.seq:
6: /cgn2_6/ptodata/1/pubpna/US10B_PUBCOMB.seq:
7: /cgn2_6/ptodata/1/pubpna/US10C_PUBCOMB.seq:
8: /cgn2_6/ptodata/1/pubpna/US10D_PUBCOMB.seq:
9: /cgn2_6/ptodata/1/pubpna/US10E_PUBCOMB.seq:
10: /cgn2_6/ptodata/1/pubpna/US11_PUBCOMB.seq:

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	2604	100.0	1518	8	US-10-411-910A-114
2	2604	100.0	1518	8	US-10-411-910A-221
3	1707	65.6	1494	8	US-10-411-910A-112
4	1702.5	65.4	1494	8	US-10-411-910A-219
5	1419.5	54.5	1311	8	US-10-411-910A-93
6	1419.5	54.5	1311	8	US-10-411-910A-200
7	1412	54.2	1248	8	US-10-411-910A-113

8	1412	54.2	1248	8	US-10-411-910A-220	Sequence 220, App
9	1408	54.1	1350	8	US-10-411-910A-92	Sequence 92, Appl
10	1408	54.1	1350	8	US-10-411-910A-199	Sequence 199, App
11	1232.5	47.3	1212	8	US-10-411-910A-115	Sequence 115, App
12	1225.5	47.1	1212	8	US-10-411-910A-222	Sequence 222, App
13	1016	39.0	1827	8	US-10-411-910A-104	Sequence 104, App
14	1016	39.0	1827	8	US-10-411-910A-211	Sequence 211, App
15	973.5	37.4	1746	8	US-10-411-910A-116	Sequence 116, App
16	973.5	37.4	1746	8	US-10-411-910A-223	Sequence 223, App
17	940	36.1	1668	8	US-10-411-910A-119	Sequence 119, App
18	940	36.1	1668	8	US-10-411-910A-226	Sequence 226, App
19	932.5	35.8	1170	8	US-10-411-910A-120	Sequence 120, App
20	932.5	35.8	1170	8	US-10-411-910A-227	Sequence 227, App
21	922.5	35.4	1911	8	US-10-411-910A-118	Sequence 118, App
22	922.5	35.4	1911	8	US-10-411-910A-225	Sequence 225, App
23	910	34.9	1821	8	US-10-411-910A-224	Sequence 224, App
24	909.5	34.9	1821	8	US-10-411-910A-117	Sequence 117, App
25	907	34.8	1740	7	US-10-282-122A-16510	Sequence 16510, A
26	899.5	34.5	1758	8	US-10-411-910A-102	Sequence 102, App
27	899	34.5	1758	8	US-10-411-910A-209	Sequence 209, App
28	893.5	34.3	1665	8	US-10-411-910A-229	Sequence 229, App
29	883	33.9	1665	8	US-10-411-910A-122	Sequence 122, App
30	880.5	33.8	1377	8	US-10-411-910A-121	Sequence 121, App
31	877.5	33.7	1377	8	US-10-411-910A-228	Sequence 228, App
32	859.5	33.0	1737	8	US-10-411-910A-123	Sequence 123, App
33	859.5	33.0	1737	8	US-10-411-910A-230	Sequence 230, App
34	832	32.0	1770	8	US-10-411-910A-110	Sequence 110, App
35	826	31.7	1770	8	US-10-411-910A-217	Sequence 217, App
36	814.5	31.3	1821	8	US-10-411-910A-97	Sequence 97, Appl
37	814.5	31.3	1821	8	US-10-411-910A-204	Sequence 204, App
38	799	30.7	1860	8	US-10-411-910A-231	Sequence 231, App
39	797.5	30.6	3621	8	US-10-411-910A-213	Sequence 213, App
40	789.5	30.3	1860	8	US-10-411-910A-124	Sequence 124, App
41	787	30.2	1407	8	US-10-411-910A-215	Sequence 215, App
42	786	30.2	3621	8	US-10-411-910A-106	Sequence 106, App
43	784.5	30.1	1407	8	US-10-411-910A-108	Sequence 108, App
44	780.5	30.0	1725	8	US-10-411-910A-201	Sequence 201, App
45	779.5	29.9	1455	8	US-10-411-910A-207	Sequence 207, App

ALIGNMENTS

RESULT 1
US-10-411-910A-114
; Sequence 114, Application US/10411910A
; Publication No. US20040209256A1
; GENERAL INFORMATION:
; APPLICANT: Dillion, Harrison F.
; TITLE OF INVENTION: Methods and Compositions for Evolving Hydrogenase Genes
; FILE REFERENCE: H2041203-P
; CURRENT APPLICATION NUMBER: US/10/411.910A
; CURRENT FILING DATE: 2003-04-12
; NUMBER OF SEQ ID NOS: 343
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 114
; LENGTH: 1518
; TYPE: DNA
; ORGANISM: Chlamydomonas reinhardtii
US-10-411-910A-114

Alignment Scores:
Pred. No.: 1.08e-265 Length: 1518
Score: 2604.00 Matches: 505
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 100.00% Indels: 0
DB: 8 Gaps: 0

US-10-763-712A-26 (1-505) x US-10-411-910A-114 (1-1518)

Qy 1 MetAlaLeuGlyLeuLeuAlaGluLeuArgAlaGlyGlnAlaValAlaCysAlaArg 20
Db 1 ATGGCGCTTGGTCTTCTTCGCGAGCTCGCGCGGTGAGCGGTGCGATGTGCTCGCGC 60

GenCore version 5.1.6
Copyright (c) 1993 - 2005 Compugen Ltd.

OM protein - nucleic search, using frame_plus_p2n model

Run on: December 25, 2005, 01:57:42 : Search time 269 Seconds
(without alignments)
3337.058 Million cell updates/sec

Title: US-10-763-712A-26
Perfect score: 2604
Sequence: 1 MALGLLAELRAGAVACARR.....HRAHELLTHVPGAEADA 505

Scoring table:

BLOSUM62
Xgapop 10.0, Xgapext 0.5
Ygapop 10.0, Ygapext 0.5
Fgapop 6.0, Fgapext 7.0
Delop 6.0, Delext 7.0

Searched: 1303057 seqs, 888780828 residues

Total number of hits satisfying chosen parameters: 2606114

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Command line parameters:
-MODEL=frame+ p2n.model -DEV=xlh
-Q=/cgn2_1/USPTO.spool/US10763712/runat 21122005 164206 10276/app query.fasta_1.647
-DB=Issued Patents NA -QFMT=fastcap -SUFFIX=p2n.rn1 -MINMATCH=0.1 -LOOPCU=0
-LOPEXT=0 -UNITS=bits -START=1 -END=1 -MATRIX=blosum62 -TRANS=human40.cdi
-LIST=45 -DOCALLIGN=200 -THR SCORE=pct -THR MAX=100 -THR MIN=0 -ALIGN=15
-MODE=LOCAL -OUTFMT=ptc -NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=2000000000
-USER=US10763712 @CGN 1 193 @runat 21122005 164206 10276 -NCPU=6 -ICPU=3
-NO MMAP -LARGEQUERY -NEG SCORES=0 -WAIT -DSPBLOCK=100 -LONGLOG
-DEV TIMEOUT=120 -WARN TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database : Issued Patents NA:

- 1: /cgn2_6/ptodata/1/ina/1 COMB.seq:*
- 2: /cgn2_6/ptodata/1/ina/5 COMB.seq:*
- 3: /cgn2_6/ptodata/1/ina/6A COMB.seq:*
- 4: /cgn2_6/ptodata/1/ina/6B COMB.seq:*
- 5: /cgn2_6/ptodata/1/ina/H COMB.seq:*
- 6: /cgn2_6/ptodata/1/ina/PCFUS COMB.seq:*
- 7: /cgn2_6/ptodata/1/ina/PP COMB.seq:*
- 8: /cgn2_6/ptodata/1/ina/RE COMB.seq:*
- 9: /cgn2_6/ptodata/1/ina/backfiles1.seq:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	1710	65.7	2399	3	US-10-077-699C-8
2	1427	54.8	2421	3	US-10-077-699C-9
3	1422	54.6	5208	3	US-10-077-699C-2
4	1401	53.8	2636	3	US-10-077-699C-7
5	1234.5	47.4	3265	3	US-10-077-699C-3
6	1130	43.4	5001	3	US-10-077-699C-1
7	478.5	18.4	2496	3	US-09-799-451-872
8	417.5	16.0	1885	3	US-09-270-767-13811
9	305.5	11.7	1791	3	US-09-248-796A-1258

10	274.5	10.5	1476	3	US-09-614-221A-377	Sequence 377, App
11	162.5	6.2	2483	3	US-10-104-047-1799	Sequence 1799, Ap
c 12	135	5.2	53500	3	US-09-266-965-76	Sequence 76, Appl
c 13	130	5.0	2028	3	US-09-252-991A-13079	Sequence 13079, A
c 14	130	5.0	3525	3	US-09-252-991A-12708	Sequence 12708, A
c 15	130	5.0	3906	3	US-09-252-991A-13251	Sequence 13251, A
c 16	129	5.0	4473	3	US-09-799-451-118	Sequence 118, App
17	127	4.9	80161	3	US-09-036-987A-1	Sequence 1, Appli
18	127	4.9	80161	3	US-09-370-700-1	Sequence 1, Appli
19	127	4.9	80161	3	US-09-603-207-1	Sequence 1, Appli
c 20	126.5	4.9	1470	3	US-09-252-991A-9149	Sequence 9149, Ap
c 21	126	4.8	3933	3	US-09-949-016-1465	Sequence 1465, Ap
c 22	126	4.8	25762	3	US-09-949-016-13207	Sequence 13207, A
23	125.5	4.8	11220	3	US-09-105-537-32	Sequence 32, Appl
24	125.5	4.8	36778	3	US-09-105-537-5	Sequence 5, Appli
25	125.5	4.8	38506	3	US-09-320-878-19	Sequence 19, Appl
26	125.5	4.8	38506	3	US-09-141-908-1	Sequence 1, Appli
27	125.5	4.8	38506	3	US-09-657-440-19	Sequence 19, Appl
28	125.5	4.8	38506	3	US-09-793-708-19	Sequence 19, Appl
c 29	125	4.8	3984	3	US-09-016-434-1199	Sequence 1199, Ap
30	125	4.8	11812	3	US-09-902-540-1041	Sequence 1041, Ap
31	124.5	4.8	4403765	3	US-09-103-840A-2	Sequence 2, Appli
32	124.5	4.8	4411529	3	US-09-103-840A-1	Sequence 1, Appli
c 33	124	4.8	3079	3	US-09-643-597-116	Sequence 116, App
c 34	124	4.8	3079	3	US-09-480-884A-116	Sequence 116, App
c 35	124	4.8	3079	3	US-09-542-615A-116	Sequence 116, App
c 36	124	4.8	3079	3	US-09-606-421B-116	Sequence 116, App
c 37	124	4.8	3079	3	US-09-221-107-116	Sequence 116, App
c 38	124	4.8	3079	3	US-09-466-396A-116	Sequence 116, App
c 39	124	4.8	3079	3	US-09-476-496A-116	Sequence 116, App
c 40	124	4.8	3079	3	US-09-630-940B-116	Sequence 116, App
c 41	124	4.8	3079	3	US-09-285-479-116	Sequence 116, App
c 42	124	4.8	3079	3	US-10-007-700-116	Sequence 116, App
c 43	123.5	4.7	16833	3	US-09-902-540-1235	Sequence 1235, Ap
c 44	122	4.7	27903	3	US-09-902-540-5492	Sequence 5492, Ap
c 45	121.5	4.7	1461	3	US-09-902-540-5492	Sequence 5492, Ap

ALIGNMENTS

RESULT 1

US-10-077-699C-8
; Sequence 8, Application US/10077699C
; Patent No. 6858718
; GENERAL INFORMATION:
; APPLICANT: Happe, Thomas
; TITLE OF INVENTION: Hydrogen Production
; FILE REFERENCE: 01MELI
; CURRENT APPLICATION NUMBER: US/10/077,699C
; CURRENT FILING DATE: 2002-02-15
; NUMBER OF SEQ ID NOS: 9
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 8
; LENGTH: 2399
; TYPE: RNA
; ORGANISM: Chlamydomonas reinhardtii
US-10-077-699C-8

Alignment Scores:

Pred. No.: 4.24e-163 Length: 2399
Score: 1710.00 Matches: 339
Percent Similarity: 76.80% Conservatives: 55
Best Local Similarity: 66.08% Mismatches: 77
Query Match: 65.67% Indels: 42
DB: 3 Gaps: 6

US-10-763-712A-26 (1-505) x US-10-077-699C-8 (1-2399)

Qy	14	AlaValAlaCysAlaArgArgThrAsnAlaProAlaHisProAlaAlaValValProCys	33
Db	151	GGCACAAGAUCCGCGCGUGUGAAGCCCGCGCGUGUCUAUCCGCGCAGCU	210
Qy	34	-LeuProSerArgAlaGlyLysPheAsnLeuSerGlnLysValProSerSerGlnSe	53

GenCore version 5.1.6
Copyright (c) 1993 - 2005 Compugen Ltd.

OM protein - nucleic search, using frame_plus_p2n model

Run on: December 25, 2005, 01:29:08 ; Search time 5783 Seconds
(without alignments)
4085.678 Million cell updates/sec

Title: US-10-763-712a-26
Perfect score: 2604
Sequence: 1 MALGALLAELRAGQAVACARR.....HRAHELLTHYVPGAREADA 505

Scoring table: BLOSUM62
Xgapop 10.0 , Xgapext 0.5
Ygapop 10.0 , Ygapext 0.5
Fgapop 6.0 , Fgapext 7.0
Delop 6.0 , Delext 7.0

Searched: 41078325 seqs, 23393541228 residues

Total number of hits satisfying chosen parameters: 82156650

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Command line parameters:
-MODEL=frame+ p2n.model -DEV=xlh
-O=/cgn2.1/USPTO.spool/US10763712/runat_21122005.164206.10264/app.query.fasta_1.647
-DB=EST -QFMT=fastap -SUFFIX=p2n.rst -MINMATCH=0.1 -LOOFC=0 -LOOPEXT=0
-UNITS=bits -START=1 -END=1 -MATRIX=blosum62 -TRANS=human40.cdi -LIST=45
-DOCALIGN=200 -THR SCORE=pct -THR MAX=100 -THR MIN=0 -ALIGN=15 -MODE=LOCAL
-OUTFMT=ptc -NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=200000000
-USER=US10763712 @CGN 1 1 5315 @runat_21122005.164206.10264 -NCPU=6 -ICPU=3
-NO_MMAP -LARGEQUERY -NEG SCORES=0 -WAIT -DSPBLOCK=100 -LONGLOG
-DEV TIMEOUT=120 -WARN TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database : EST:*

1: gb_est1:*
2: gb_est2:*
3: gb_est3:*
4: gb_hic:*
5: gb_est4:*
6: gb_est5:*
7: gb_est6:*
8: gb_est7:*
9: gb_gss1:*
10: gb_gss2:*
11: gb_gss3:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1112	42.7	578	3	BI726441
2	930	35.7	658	3	BI529458
3	919	35.3	661	1	AW721369
4	906	34.8	549	2	BE337478
5	861.5	33.1	617	3	BI999849
6	843	32.4	505	2	BE452899
7	796	30.6	644	3	BM003317

8	760	29.2	505	2	BE238123
9	749	28.8	583	3	BI995712
c 10	736.5	28.3	1033	7	CV210633
c 11	655.5	25.2	973	7	CV209966
c 12	617.5	23.7	883	7	CV209967
c 13	583	22.4	836	7	CV210634
c 14	556.5	21.4	558	1	AV189547
c 15	536	20.6	385	2	BE725875
c 16	520.5	20.0	1050	10	CW918425
c 17	513	19.7	398	3	BP095261
c 18	508	19.5	1431	10	AY410480
c 19	508	19.5	1999	4	BC020468
c 20	508	19.5	2043	4	AK018548
c 21	506	19.4	2532	4	AK043694
c 22	505	19.4	2003	4	AK014038
c 23	502	19.3	457	1	AV624332
c 24	484	18.6	599	3	BI716925
c 25	479.5	18.4	2105	4	CR857325
c 26	478.5	18.4	1431	10	AY410478
c 27	461	17.7	1118	10	CW934222
c 28	438.5	16.8	1683	4	CNS0A3NF
c 29	427.5	16.4	1905	4	CR595518
c 30	418.5	16.1	2037	4	AK013432
c 31	418.5	16.1	2308	4	AK081305
c 32	418.5	16.1	4392	4	AK076232
c 33	415.5	16.0	603	7	CV216048
c 34	409	15.7	657	3	BI717904
c 35	394.5	15.1	601	7	CV216049
c 36	385.5	14.8	1265	4	CR601772
c 37	385.5	14.8	1469	4	CR611448
c 38	385.5	14.8	1516	4	CR603660
c 39	385.5	14.8	1524	4	CR597828
c 40	385.5	14.8	1539	4	CR606462
c 41	385.5	14.8	1542	4	CR595982
c 42	385.5	14.8	1574	4	CR598397
c 43	385	14.8	548	3	BI532215
c 44	384	14.7	1196	4	CR619657
c 45	379	14.6	343	6	CF562921

ALIGNMENTS

RESULT 1

BI726441

LOCUS

DEFINITION

ACCESSION

VERSION

KEYWORDS

SOURCE

ORGANISM

REFERENCE

AUTHORS

TITLE

JOURNAL

COMMENT

BI726441 658 bp mRNA linear EST 19-SEP-2001
1031085C09.y1 C. reinhardtii CC-1690, Stress II (normalized),
Lambda zap II Chlamydomonas reinhardtii cDNA, mRNA sequence.

BI726441.1 GI:15702136

EST.

Chlamydomonas reinhardtii

Chlamydomonas reinhardtii

Eukaryota: Viridiplantae;

Chlamydomonadales; Chlamydomonadaceae; Chlamydomonas.

1 (bases 1 to 658)

Grossman, A., Chang, C.-W., Davies, J., Harris, E., Hauser, C.,

Lefebvre, P., McDermott, J.P., Shrager, J., Silflow, C. and Stern, D.

Analyses of the Chlamydomonas reinhardtii Genome: A Model.

Uncellular System for Analyzing Gene Function and Regulation in

Vascular Plants. Project: 1031

Unpublished (2001)

Contact: Charles Hauser

DCMB Box 91000

Duke University

Durham, NC 27708-1000

Tel: 919 613 8159

Fax: 919 613 8177

Email: chauser@duke.edu.

Location/Qualifiers

1. 658

/organism="Chlamydomonas reinhardtii"

/mol_type="mRNA"

/strain="CC-1690 wild type mt+ 21gr"

FEATURES
source

GenCore version 5.1.6
Copyright (c) 1993 - 2005 Compugen Ltd.

OM protein - nucleic search, using frame_plus_p2n model

Run on: December 24, 2005, 22:18:19 ; Search time 871 Seconds

(without alignments)
3864.144 Million cell updates/sec

Title: US-10-763-712A-26

Perfect score: 2604

Sequence: 1 MAUGLIALRAGQAVACARR.....HRAHELHLLHYVPGAEADA 505

Scoring table:

BLOSUM62
Xgapop 10.0 , Xgapext 0.5
Ygapop 10.0 , Ygapext 0.5
Fgapop 6.0 , Fgapext 7.0
Delop 6.0 , Delext 7.0

Searched: 4996997 seqs, 3332346308 residues

Total number of hits satisfying chosen parameters: 9993994

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Command line parameters:

-MODEL=frame+ p2n.model -DEV=xlh
-O=/cpn2_1/USPTO.epool/US10763712/runat_21122005_164206_10240/app_query.fasta_1.647
-DB=Geneseq -QFMT=fastcap -SUFFIX=p2n.rng -MINMATCH=0.1 -LOPCL=0 -LOPEXT=0
-UNITS=bits -START=1 -END=-1 -MATRIX=blomsum62 -TRANS=numa40.cdi -LIST=45
-DOCALIGN=200 -THRS=score=pct -THR MAX=100 -THR MIN=0 -ALIGN=15 -MODE=LOCAL
-OUTFMT=ptc -NORM=ext -HEADSIZE=500 -MINLEN=0 -MAXLEN=2000000000
-USER=US10763712 @CGN 1.1 727 @runat_21122005_164206_10240 -NCPU=6 -ICPU=3
-NO MMAP -LARGEQUERY -NEG_SCORES=0 -WAIT -DSPLOCK=100 -LONGLOG
-DEV_TIMEOUT=120 -WARN_TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAEXT=0.5 -FGAPOP=6
-FGAEXT=7 -YGAPOP=10 -YGAEXT=0.5 -DELOP=6 -DELEXT=7

Database : N Geneseq 21.*

1: Geneseqn1980s.*
2: Geneseqn1990s.*
3: Geneseqn2000s.*
4: Geneseqn2001as.*
5: Geneseqn2001bs.*
6: Geneseqn2002as.*
7: Geneseqn2002bs.*
8: Geneseqn2003as.*
9: Geneseqn2003bs.*
10: Geneseqn2003cs.*
11: Geneseqn2003ds.*
12: Geneseqn2004as.*
13: Geneseqn2004bs.*
14: Geneseqn2005s.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	1710	65.7	2399	14	ADY27752 Chlamydom
2	1707	65.6	1494	14	Aeb73387 cDNA enco
3	1427	54.8	2421	14	ADY27753 Chlorella
4	1422	54.6	5200	14	ADY27766 Chlamydom

5	1422	54.6	5208	14	ADY27746	Ady27746 Chlamydom
6	1419.5	54.5	1311	14	AEB73392	Aeb73392 DNA enco
7	1408	54.1	1350	14	AEB73391	Aeb73391 DNA enco
8	1401	53.8	2636	14	ADY27751	Ady27751 Scenedesm
9	1234.5	47.4	3265	14	ADY27747	Ady27747 Chlorella
10	1130	43.4	5001	14	ADY27745	Ady27745 Scenedesm
11	907	34.8	1740	8	ACA28640	ACA28640 Prokaryot
12	791	30.4	2347	10	ADC59543	ADC59543 Clostridi
13	791	30.4	2347	13	ADP90306	ADP90306 Clostridi
14	784.5	30.1	1544	14	ADZ39228	Adz39228 Trichomon
15	754.5	29.0	1900	2	AAZ25197	AAZ25197 Nucleotid
16	750.5	28.8	1725	14	AEB73388	Ady277388 cDNA enco
17	742.5	28.5	1737	2	AAZ25199	AAZ25199 Clostridi
18	740.5	28.4	1737	2	AAZ25198	AAZ25198 Clostridi
19	716	27.5	1749	8	ACA27971	ACA27971 Prokaryot
20	696.5	26.7	1964	14	ADZ39223	Adz39223 Deulfovi
21	695	26.7	1800	14	ADZ39224	Adz39224 Deulfovi
22	672.5	25.8	1265	14	AEB73389	Aeb73389 cDNA enco
23	478.5	18.4	2101	4	AAH24245	Aah24245 Human oxi
24	478.5	18.4	2496	6	ABZ11990	Abz11990 Human pol
25	478.5	18.4	2496	12	ADMA4508	Adm44508 Novel hum
26	474	18.2	1407	13	ADT91276	Adt91276 Entamoeba
27	472	18.1	1407	14	AEB73390	Aeb73390 DNA enco
28	448	17.2	2113	4	AHC91328	Aac91328 Human pol
29	409	15.7	1401	13	ADT91275	Adt91275 Entamoeba
30	401.5	15.4	1371	10	ADC29959	Adc29959 Human nov
31	400.5	15.4	1453	8	ACC47252	Acc47252 Human SCA
32	400.5	15.4	1465	8	ACC44314	Acc44314 Gene enco
33	398	15.3	3219	4	ABL25132	Ab125132 Drosophil
34	391.5	15.0	1424	4	ABL25133	Ab125133 Drosophil
35	385.5	14.8	1562	4	AAH13714	Aah13714 Human cDN
36	385.5	14.8	1577	5	AAH27127	Aah27127 Human NAD
37	385.5	14.8	1606	13	ADR25541	Adr25541 Breast ca
38	380.5	14.6	1760	10	ADF76547	Adf76547 Novel hum
39	380.5	14.5	1760	13	ACN40308	Acn40308 Tumour-as
40	371	14.2	1467	14	AEAL9500	Aeal9500 Novel hum
41	367.5	14.1	1776	8	ABT19372	Abt19372 Aspergill
42	362.5	13.9	1592	11	ADM06990	Adm06990 Aspergill
43	362.5	13.9	1725	11	ADM07098	Adm07098 Aspergill
44	361.5	13.9	1735	11	ADM06989	Adm06989 Aspergill
45	361.5	13.9	1821	8	ABT21192	Abt21192 Aspergill

ALIGNMENTS

RESULT 1

ADY27752

ID ADY27752 standard; mRNA; 2399 BP.

XX AC ADY27752;

XX DT 05-MAY-2005 (first entry)

XX DE Chlamydomonas reinhardtii iron hydrogenase hyda mRNA, SEQ ID NO:8.

XX KW Iron hydrogenase; Hyda; plant; gene; ss.

XX OS Chlamydomonas reinhardtii; strain 137c(mt+).

XX FH Key Location/Qualifiers

XX FT CDS 159..1652

XX FT /*tag= b

XX FT /product= "Iron hydrogenase (Hyda)"

XX FT transit_peptide 159..326

XX FT /*tag= a

XX FT /label= Chloroplast_ströma_transit_peptide

XX FT mat_peptide 327..1649

XX FT /*tag= c

XX FT /product= "Mature iron hydrogenase"

XX PN US6859718-B1.

XX XX

PD 22-FEB-2005.

GenCore version 5.1.6
Copyright (c) 1993 - 2005 Compugen Ltd.

OM protein - nucleic search, using frame plus p2n model
(without alignments)

Run on: December 25, 2005, 01:27:12 ; Search time 7549 Seconds
3802.615 Million cell updates/sec

Title: US-10-763-712A-26

Perfect score: 2604

Sequence: 1 MALGLLAEIRAGQAVACARR.....HRAELHLLTHYVPGAEADA 505

Scoring table:

BLOSUM62
Xgapop 10.0, Xgapext 0.5
Ygapop 10.0, Ygapext 0.5
Fgapop 6.0, Fgapext 7.0
Delop 6.0, Delext 7.0

Searched: 5883141 seqs, 28421725653 residues

Total number of hits satisfying chosen parameters: 11766282

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Command line parameters: -DB=frame+ p2n model -DEV=xlh
-Q=/cgn2_1/USPTO spo01/US10763712/runat_21122005_164206_10250/app_query.fasta_1.647
-DB=GenEmbl -QMT=fastcap -SUFFIX=p2n.rge -MINMATCH=0.1 -LOOPCL=0 -LOOPEXT=0
-UNITS=bits -START=1 -END=1 -MATRIX=biosum62 -TRANS=human40.cdi -LIST=45
-DOCALIGN=200 -THR SCORE=pct -THR MAX=100 -THR MIN=0 -ALIGN=15 -MODE=LOCAL
-OUTFMT=ptc -NORM=ext -HEAPSIZ=500 -MINLEN=0 -MAXLEN=200000000
-USER=US10763712 @CGN 1.1 4939 @runat_21122005_164206_10250 -NCPU=6 -ICPU=3
-NO_MMAP -LARGQUERY -NEG_SCORES=0 -WAIT_DSPLCK=100 -LONGLOG
-DEV_TIMEOUT=120 -WARN_TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database :

GenEmbl:
1: gb.ba.*
2: gb.in.*
3: gb.env.*
4: gb.om.*
5: gb.ov.*
6: gb.pat.*
7: gb.ph.*
8: gb.pr.*
9: gb.ro.*
10: gb.sts.*
11: gb.sy.*
12: gb.un.*
13: gb.vi.*
14: gb.htg.*
15: gb.pl.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	2604	100.0	2560	15	AY055756 Chlamydom
2	2542	97.6	1518	15	AY436606 Chlamydom
3	2087	80.1	4622	15	AY090770 Chlamydom

	4	2035.5	78.2	3584	15	AY436607
5	1710	65.7	2399	6	AR641147	AY436607 Chlamydom
6	1710	65.7	2399	15	CRE012098	AR641147 Sequence
7	1710	65.7	2454	15	AY055755	AJ012098 Chlamydom
8	1710	65.7	2522	15	AF289201	AY055755 Chlamydom
9	1427	54.8	2420	15	CFU298228	AF289201 Chlamydom
10	1427	54.8	2421	6	AR641148	AJ298228 Chlorella
11	1427	54.6	5208	6	AR641144	AR641148 Sequence
12	1427	54.6	5208	15	CRE308413	AR641144 Sequence
13	1401	53.8	2636	6	AR641146	AJ308413 Chlamydom
14	1329	51.0	1853	15	AY578072	AR641146 Sequence
15	1234.5	47.4	3265	6	AR641145	AY578072 Chlamydom
16	1234.5	47.4	3265	15	CFU298227	AR641145 Sequence
17	1232.5	47.3	1607	15	AF276706	AJ298227 Chlorella
18	1130	43.4	5001	6	AR641143	AF276706 Scenedes
19	1130	43.4	5001	15	SOB271546	AR641143 Sequence
20	1016	39.0	15158	1	AE001705	AJ271546 Scenedes
21	975.5	37.5	110000	1	AP006840_34	AE001705 Thermotog
22	973.5	37.4	11577	1	AE013056	Continuation (35 o
23	940	36.1	1761	15	AF446076	AE013056 Thermotog
24	932.5	35.8	1337	15	AF516683	AF446076 Piromyces
25	922.5	35.4	2467	15	AY033895	AF516683 Neocallim
26	902.5	34.7	2238	1	AF148212	AY033895 Neocallim
27	899.5	34.5	5251	1	DFU07229	AF148212 Clostridi
28	892.5	34.3	304143	1	AE016926	DFU07229 Desulfovib
29	881	33.8	66135	1	AP006840_35	AE016926 Bacteroid
30	872	33.5	110000	1	CP000027_01	Continuation (36 o
31	862	33.1	110000	1	CR522870_04	Continuation (2 of
32	862	33.1	110000	1	CR522870_05	Continuation (5 of
33	859.5	33.0	16133	1	EAC312124	Continuation (6 of
34	832	32.0	1848	2	AF262401	AJ312124 Eubacteri
35	814.5	31.3	1940	1	DVHYDC	AF262401 Trichomon
36	814.5	31.3	110000	2	AE017314_3	X57838 D.vulgaris
37	802	30.8	3933	2	AY608627_	Continuation (4 of
38	791	30.4	2347	1	AB159510	AY608627 Nyctother
39	789.5	30.3	349315	2	BX572593	AB159510 Clostridi
40	786	30.2	3625	2	NH572593	BX572593 Rhodosphe
41	784.5	30.1	1544	2	TVU19897	Y19897 Rhodosphe
42	784.5	30.1	1544	6	GS073824	Y19897 Trichomonas
43	774	29.7	2585	1	AF120457	CS073824 Sequence
44	767	29.5	2267	1	CAU09760	AF120457 Megasphe
45	750.5	28.8	2310	1	CL0HDGI	U09760 Clostridium
						M81737 Clostridium

ALIGNMENTS

RESULT 1	AY055756	2560 bp	mRNA	linear	PLN 24-MAR-2005
LOCUS	AY055756	Chlamydomonas reinhardtii iron-hydrogenase Hyda2 (hyda2) mRNA,			
DEFINITION	AY055756	complete cds.			
ACCESSION	AY055756	GI:18026271			
VERSION	AY055756				
KEYWORDS	Chlamydomonas reinhardtii				
SOURCE	Chlamydomonas reinhardtii				
ORGANISM	Eukaryota: Viridiplantae; Chlorophyta: Chlorophyceae;				
REFERENCE	Forestier, M., King, P., Zhang, L., Posewitz, M., Schwarzer, S.,				
AUTHORS	Happe, T., Ghirardi, M.L. and Seibert, M.				
TITLE	Expression of two [Fe]-hydrogenases in Chlamydomonas reinhardtii				
JOURNAL	under anaerobic conditions				
PUBMED	Eur. J. Biochem. 270 (13), 2750-2758 (2003)				
REFERENCE	12823345				
AUTHORS	2 (bases 1 to 2560)				
TITLE	Forestier, M., Zhang, L., Plummer, S., Ahmann, D., Seibert, M. and				
JOURNAL	Ghirardi, M.L.				
REFERENCE	Two putative Fe-only hydrogenases cloned from Chlamydomonas				
AUTHORS	reinhartii are coexpressed in cells undergoing anaerobiosis				
TITLE	Unpublished				
JOURNAL	3 (bases 1 to 2560)				
REFERENCE	Forestier, M., Zhang, L., Plummer, S., Ahmann, D., Seibert, M. and				
AUTHORS					

GenCore version 5.1.6
Copyright (c) 1993 - 2005 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: December 21, 2005, 23:02:58 ; Search time 12 Seconds
(without alignments)
300.143 Million cell updates/sec

Title: US-10-763-712A-26
Perfect score: 2604
Sequence: 1 MALGILAEIRAGQAVACARR.....HRAHELLTHYVPGAREADA 505

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 53982 seqs, 7132107 residues

number of hits satisfying chosen parameters: 53982

DB seq length: 0
DB seq length: 2000000000
Processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

- Published Applications AA New:*
- 1: /cgn2_6/ptodata/2/pubpaa/US08_NEW_PUB.pep.*
 - 2: /cgn2_6/ptodata/2/pubpaa/US06_NEW_PUB.pep.*
 - 3: /cgn2_6/ptodata/2/pubpaa/US07_NEW_PUB.pep.*
 - 4: /cgn2_6/ptodata/2/pubpaa/PCT_NEW_PUB.pep.*
 - 5: /cgn2_6/ptodata/2/pubpaa/US05_NEW_PUB.pep.*
 - 6: /cgn2_6/ptodata/2/pubpaa/US10_NEW_PUB.pep.*
 - 7: /cgn2_6/ptodata/2/pubpaa/US11_NEW_PUB.pep.*
 - 8: /cgn2_6/ptodata/2/pubpaa/US60_NEW_PUB.pep.*

Red. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Query				ID	Description
Score	Match	Length	DB		
2604	100.0	505	6	US-10-763-712A-26	Sequence 26, Appl
2598	99.8	505	6	US-10-763-712A-123	Sequence 123, Appl
2549	97.9	505	6	US-10-763-712A-88	Sequence 88, Appl
2542	97.6	505	6	US-10-763-712A-89	Sequence 89, Appl
1702.5	65.4	497	6	US-10-763-712A-24	Sequence 24, Appl
1702.5	65.4	497	6	US-10-763-712A-91	Sequence 91, Appl
1419.5	54.5	436	6	US-10-763-712A-5	Sequence 5, Appl
1412	54.2	415	6	US-10-763-712A-25	Sequence 25, Appl
1408	54.1	449	6	US-10-763-712A-83	Sequence 83, Appl
1225.5	47.1	403	6	US-10-763-712A-27	Sequence 27, Appl
1016	39.0	608	6	US-10-763-712A-16	Sequence 16, Appl
1016	39.0	608	6	US-10-763-712A-90	Sequence 90, Appl
973.5	37.4	581	6	US-10-763-712A-28	Sequence 28, Appl
973.5	37.4	581	6	US-10-763-712A-92	Sequence 92, Appl
940	36.1	555	6	US-10-763-712A-30	Sequence 30, Appl
932.5	35.8	389	6	US-10-763-712A-31	Sequence 31, Appl
922.5	35.4	636	6	US-10-763-712A-29	Sequence 29, Appl
922.5	35.4	636	6	US-10-763-712A-93	Sequence 93, Appl
909.5	34.9	606	6	US-10-763-712A-86	Sequence 86, Appl
902.5	34.7	579	6	US-10-763-712A-94	Sequence 94, Appl
896.5	34.4	585	6	US-10-763-712A-14	Sequence 14, Appl
896.5	34.4	585	6	US-10-763-712A-96	Sequence 96, Appl
884.5	34.0	588	6	US-10-763-712A-95	Sequence 95, Appl
881	33.8	554	6	US-10-763-712A-33	Sequence 33, Appl
873	33.5	458	6	US-10-763-712A-32	Sequence 32, Appl

ALIGNMENTS

RESULT 1

US-10-763-712A-26
; Sequence 26, Application US/10763712A
; Publication No. US20050266541A1
; GENERAL INFORMATION:
; APPLICANT: Solazyme, Inc.
; APPLICANT: Dillon, Harrison F.
; TITLE OF INVENTION: Methods and Compositions for Evolving Microbial Hydrogen
; TITLE OF INVENTION: Production
; FILE REFERENCE: H2042101-CIP
; CURRENT APPLICATION NUMBER: US/10/763,712A
; PRIOR FILING DATE: 2004-01-21
; PRIOR APPLICATION NUMBER: US 10/287,750
; PRIOR FILING DATE: 2002-11-04
; PRIOR APPLICATION NUMBER: US 10/411,910
; PRIOR FILING DATE: 2003-04-12
; PRIOR APPLICATION NUMBER: US 60/500,032
; PRIOR FILING DATE: 2003-09-03
; NUMBER OF SEQ ID NOS: 184
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 26
; LENGTH: 505
; TYPE: PRT
; ORGANISM: Chlamydomonas reinhardtii
US-10-763-712A-26

Query Match	100.0%	Score 2604;	DB 6;	Length 505;
Best Local Similarity	100.0%	Pred. No. 3.4e-200;	Mismatches 0;	Indels 0;
Matches 505;	Conservative 0;			
Qy	1	MALGILAEIRAGQAVACARRNAPAHVAVPCLPSRAGKFFNLQKVPSSQSARGSTIR	60	
Db	1	MALGILAEIRAGQAVACARRNAPAHVAVPCLPSRAGKFFNLQKVPSSQSARGSTIR	60	
Qy	61	VAATATDAPVHWKLAELDKPKGGRKVLIAQVAPVAVTAESFGLAPGAVSPGKLAT	120	
Db	61	VAATATDAPVHWKLAELDKPKGGRKVLIAQVAPVAVTAESFGLAPGAVSPGKLAT	120	
Qy	121	GLRALGPQVQVDTTLFAADLTIMEEGTELLHRLKEHLEAHPHSDEPLPMTSCCPGVNMM	180	
Db	121	GLRALGPQVQVDTTLFAADLTIMEEGTELLHRLKEHLEAHPHSDEPLPMTSCCPGVNMM	180	
Qy	181	EKSYPELLIPFYSSCKSPQMMGMVTKTYLSEKQIGIPAKDI VMVSVMPVCVRKQGEADREWF	240	
Db	181	EKSYPELLIPFYSSCKSPQMMGMVTKTYLSEKQIGIPAKDI VMVSVMPVCVRKQGEADREWF	240	
Qy	241	CVSFPGRVDRVITTTABELGNI FKERGINLPDPSDWDQPLGLSGAGVLFGTGGVME	300	

GenCore version 5.1.6
Copyright (c) 1993 - 2005 Compugen Ltd.

OM protein - protein search, using sw model

Run on: December 21, 2005, 22:59:43 ; Search time 165 Seconds
(without alignments)
1278.811 Million cell updates/sec

Title: US-10-763-712A-26

Perfect score: 2604

Sequence: 1 MAUGLLAELRAGQAVACARR.....HRAHELLLTHYVPGABADA 505

Scoring table:

BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 1867569 seqs, 417829326 residues

Total number of hits satisfying chosen parameters: 1867569

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

Published Applications AA_Main:
1: /cgn2_6/ptodata/1/pubpaa/US07_PUBCOMB.pep.*
2: /cgn2_6/ptodata/1/pubpaa/US08_PUBCOMB.pep.*
3: /cgn2_6/ptodata/1/pubpaa/US09_PUBCOMB.pep.*
4: /cgn2_6/ptodata/1/pubpaa/US10A_PUBCOMB.pep.*
5: /cgn2_6/ptodata/1/pubpaa/US10B_PUBCOMB.pep.*
6: /cgn2_6/ptodata/1/pubpaa/US11_PUBCOMB.pep.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	2604	100.0	505	4 US-10-411-910A-26	Sequence 26, Appl
2	1702.5	65.4	497	4 US-10-411-910A-24	Sequence 24, Appl
3	1419.5	54.5	436	4 US-10-411-910A-5	Sequence 5, Appl
4	1412	54.2	415	4 US-10-411-910A-25	Sequence 25, Appl
5	1408	54.1	449	4 US-10-411-910A-83	Sequence 83, Appl
6	1235.5	47.1	403	4 US-10-411-910A-27	Sequence 27, Appl
7	1016	39.0	608	4 US-10-411-910A-16	Sequence 16, Appl
8	973.5	37.4	581	4 US-10-411-910A-28	Sequence 28, Appl
9	940	36.1	555	4 US-10-411-910A-30	Sequence 30, Appl
10	932.5	35.8	389	4 US-10-411-910A-31	Sequence 31, Appl
11	922.5	35.4	636	4 US-10-411-910A-29	Sequence 29, Appl
12	909.5	34.9	606	4 US-10-411-910A-86	Sequence 86, Appl
13	907	34.8	580	4 US-10-411-910A-26	Sequence 26, Appl
14	896.5	34.4	585	4 US-10-411-910A-14	Sequence 14, Appl
15	881	33.8	554	4 US-10-411-910A-33	Sequence 33, Appl
16	873	33.5	458	4 US-10-411-910A-32	Sequence 32, Appl
17	859.5	33.0	578	4 US-10-411-910A-34	Sequence 34, Appl
18	818.5	31.4	589	4 US-10-411-910A-22	Sequence 22, Appl
19	814.5	31.3	606	4 US-10-411-910A-9	Sequence 9, Appl
20	789.5	30.3	619	4 US-10-411-910A-35	Sequence 35, Appl
21	786	30.2	1206	4 US-10-411-910A-18	Sequence 18, Appl
22	784.5	30.1	468	4 US-10-411-910A-20	Sequence 20, Appl
23	774	29.7	484	4 US-10-411-910A-12	Sequence 12, Appl
24	769	29.5	644	4 US-10-411-910A-36	Sequence 36, Appl
25	767	29.5	574	4 US-10-411-910A-6	Sequence 6, Appl
26	750.5	28.8	574	4 US-10-411-910A-1	Sequence 1, Appl
27	741	28.5	572	4 US-10-411-910A-10	Sequence 10, Appl

28 741 28.5 572 4 US-10-411-910A-37 Sequence 37, Appl
29 739 28.4 572 4 US-10-411-910A-11 Sequence 11, Appl
30 716 27.5 582 4 US-10-282-122A-52025 Sequence 52025, A
31 716 27.5 582 4 US-10-411-910A-23 Sequence 23, Appl
32 693.5 26.6 583 4 US-10-411-910A-38 Sequence 38, Appl
33 689 26.5 421 4 US-10-411-910A-15 Sequence 15, Appl
34 660 25.3 421 4 US-10-411-910A-2 Sequence 2, Appl
35 659 25.3 439 4 US-10-411-910A-39 Sequence 39, Appl
36 639.5 24.6 449 4 US-10-411-910A-7 Sequence 7, Appl
37 636 24.4 421 4 US-10-411-910A-13 Sequence 13, Appl
38 635.5 24.4 421 4 US-10-411-910A-41 Sequence 41, Appl
39 633.5 24.3 421 4 US-10-411-910A-85 Sequence 85, Appl
40 624.5 24.0 410 4 US-10-411-910A-17 Sequence 17, Appl
41 572 22.0 645 4 US-10-411-910A-40 Sequence 40, Appl
42 558.5 21.4 379 4 US-10-411-910A-19 Sequence 19, Appl
43 508 19.5 467 4 US-10-411-910A-45 Sequence 45, Appl
44 492 18.9 369 4 US-10-411-910A-45 Sequence 45, Appl
45 483.5 18.6 369 4 US-10-411-910A-43 Sequence 43, Appl

ALIGNMENTS

RESULT 1

US-10-411-910A-26

; Sequence 26, Application US/10411910A

; Publication No. US20040209256A1

; GENERAL INFORMATION:

; APPLICANT: Dillion, Harrison F.

; TITLE OF INVENTION: Methods and Compositions for Evolving Hydrogenase Genes

; FILE REFERENCE: H2041203-P

; CURRENT APPLICATION NUMBER: US/10/411,910A

; CURRENT FILING DATE: 2003-04-12

; NUMBER OF SEQ ID NOS: 343

; SOFTWARE: PatentIn version 3.2

; SEQ ID NO 26

; LENGTH: 505

; TYPE: PRT

; ORGANISM: Chlamydomonas reinhardtii

US-10-411-910A-26

Query Match 100.0%; Score 2604; DB 4; Length 505;

Best Local Similarity 100.0%; Pred. No. 8.8e-230;

Matches 505; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 MAUGLLAELRAGQAVACARRTNAHPAAVVPCLPSRAGKFFNLQKVPSSQSARGSTIR 60
Db 1 MAUGLLAELRAGQAVACARRTNAHPAAVVPCLPSRAGKFFNLQKVPSSQSARGSTIR 60
Qy 61 VAATATDAVPHWKLALELDKPKDGGRRKVLIAQVAPVAVRVAIAESFGLAPGAVSPGKLAT 120
Db 61 VAATATDAVPHWKLALELDKPKDGGRRKVLIAQVAPVAVRVAIAESFGLAPGAVSPGKLAT 120
Qy 121 GLRAGFDQVDTLFAADLTIMEEGTELLHRLKEHLEAHPHSDEPLPMTSCCPGWAMM 180
Db 121 GLRAGFDQVDTLFAADLTIMEEGTELLHRLKEHLEAHPHSDEPLPMTSCCPGWAMM 180
Qy 181 EKSPVLPITVSSCKSPQMMGMAMVTKYLSKQGIKPAKDIVMVSVMPVCVRQGEADREWF 240
Db 181 EKSPVLPITVSSCKSPQMMGMAMVTKYLSKQGIKPAKDIVMVSVMPVCVRQGEADREWF 240
Qy 241 CVSEPGVRDHDVHTTAEALGNIFKRGINLPDSDWDQPLGLGSGAGVLFGTGGVME 300
Db 241 CVSEPGVRDHDVHTTAEALGNIFKRGINLPDSDWDQPLGLGSGAGVLFGTGGVME 300
Qy 301 AALRTAYEIVTKPELPRLNLSEVRGLDGIKEASVTLVPAGSKFAELVAERLAHKVSEAA 360
Db 301 AALRTAYEIVTKPELPRLNLSEVRGLDGIKEASVTLVPAGSKFAELVAERLAHKVSEAA 360
Qy 361 AAEAAAAVEGAVRPPPIAYDGGQGFSTDDGKGLKLRVAVANGLNNAKKLIGMWSGEAKY 420
Db 361 AAEAAAAVEGAVRPPPIAYDGGQGFSTDDGKGLKLRVAVANGLNNAKKLIGMWSGEAKY 420

GenCore version 5.1.6
Copyright (c) 1993 - 2005 Compugen Ltd.

OM protein - protein search, using sw model

Run on: December 21, 2005, 22:59:27 ; Search time 46 Seconds
(without alignments)
907.635 Million cell updates/sec

Title: US-10-763-712A-26
Perfect score: 2604
Sequence: 1 MGLGLAELRAGQAVACARR.....HRAHELLTHYVPGARADA 505

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 572060 seqs, 82675679 residues

Total number of hits satisfying chosen parameters: 572060

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

- 1: /cgm2_6/ptodata/1/iaa/5 COMB.pap:*
- 2: /cgm2_6/ptodata/1/iaa/6 COMB.pap:*
- 3: /cgm2_6/ptodata/1/iaa/H COMB.pap:*
- 4: /cgm2_6/ptodata/1/iaa/PCTUS COMB.pap:*
- 5: /cgm2_6/ptodata/1/iaa/RE COMB.pap:*
- 6: /cgm2_6/ptodata/1/iaa/backfiles.pap:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1702.5	65.4	497	2	US-10-077-699C-5
2	1419.5	54.5	436	2	US-10-077-699C-6
3	1400.5	53.8	448	2	US-10-077-699C-4
4	417.5	16.0	484	2	US-09-270-767-45381
5	305.5	11.7	596	2	US-09-248-796A-15361
6	160	6.1	189	2	US-10-104-047-3769
7	115.5	4.4	562	2	US-09-902-540-9872
8	114.5	4.4	719	2	US-09-902-540-11526
9	113.5	4.4	502	2	US-09-252-991A-25502
10	112	4.3	574	2	US-09-949-016-5915
11	112	4.3	587	2	US-09-949-016-11334
12	110.5	4.2	1174	2	US-09-252-991A-29279
13	110.5	4.2	4545	1	US-08-804-227C-14
14	110.5	4.2	4550	1	US-08-804-227C-8
15	110.5	4.2	4550	1	US-08-804-198-2
16	109.5	4.2	850	2	US-09-252-991A-20891
17	109	4.2	688	2	US-09-477-962-96
18	106.5	4.1	440	2	US-09-252-991A-30740
19	106.5	4.1	540	2	US-09-252-991A-26410
20	106.5	4.1	3546	2	US-09-679-279-13
21	106	4.1	503	2	US-09-604-231-20
22	106	4.1	568	2	US-09-950-788B-2
23	106	4.1	568	2	US-09-950-788B-4
24	106	4.1	568	2	US-09-950-788B-7
25	106	4.1	568	2	US-09-604-231-18
26	105.5	4.1	486	2	US-09-902-540-12750
27	105	4.0	3739	2	US-09-320-878-2

28	105	4.0	3739	2	US-09-105-537-33	Sequence 33, Appli
29	105	4.0	3739	2	US-09-141-908-3	Sequence 3, Appli
30	105	4.0	3739	2	US-09-657-440-2	Sequence 2, Appli
31	105	4.0	3739	2	US-09-793-708-2	Sequence 2, Appli
32	105	4.0	11877	2	US-09-105-537-6	Sequence 6, Appli
33	104.5	4.0	3491	1	US-07-642-734C-2	Sequence 2, Appli
34	104.5	4.0	3491	2	US-08-439-009A-2	Sequence 2, Appli
35	104	4.0	4472	1	US-08-804-227C-2	Sequence 2, Appli
36	103.5	4.0	500	2	US-09-540-236-2597	Sequence 2597, Ap
37	103.5	4.0	649	2	US-09-252-991A-27657	Sequence 27657, A
38	103.5	4.0	1501	2	US-09-252-991A-20266	Sequence 20266, A
39	102	3.9	4551	2	US-09-320-878-1	Sequence 1, Appli
40	102	3.9	4551	2	US-09-141-908-2	Sequence 2, Appli
41	102	3.9	4551	2	US-09-657-440-1	Sequence 1, Appli
42	102	3.9	4551	2	US-09-793-708-1	Sequence 1, Appli
43	102	3.9	4613	2	US-09-105-537-31	Sequence 31, Appli
44	101.5	3.9	714	2	US-09-192-056-2	Sequence 2, Appli
45	101.5	3.9	898	2	US-09-252-991A-23689	Sequence 23689, A

ALIGNMENTS

RESULT 1

US-10-077-699C-5
; Sequence 5, Application US/10077699C
; Patent No. 6858718
; GENERAL INFORMATION:
; APPLICANT: Hapde, Thomas
; TITLE OF INVENTION: Hydrogen Production
; FILE REFERENCE: OIMELI
; CURRENT APPLICATION NUMBER: US/10/077.699C
; CURRENT FILING DATE: 2002-02-15
; NUMBER OF SEQ ID NOS: 9
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 5
; LENGTH: 497
; TYPE: PRT
; ORGANISM: Chlamydomonas reinhardtii
US-10-077-699C-5

Query Match 65.4%; Score 1702.5; DB 2; Length 497;
Best Local Similarity 67.7%; Pred. No. 1.9e-174;
Matches 331; Conservative 57; Mismatches 84; Indels 17; Gaps 5;

Qy	16	ACARRTNAPHAAPVAVPCLPSRAGKFFNLQKVPSSQSARGSTIRVAATATDAVPHKLA	75
Db	18	SCRARQVAPRAPLAASVTRVALA-----TLEAPARRLGNVACAAAPAEAPUSHVOQA	71
Qy	76	LEELDKPKDG-GRKVLIAQVAPVAVIAESFGLAPGAVSPGKLTGLRALGFDQVFDTL	134
Db	72	LAELAKPKDDPTRKHVCVQVAPVAVIAETLGLAPGATTPKQLAEGRLRGDFDEVFDL	131
Qy	135	PAADLTIMESGTELLHRLKEHLEHPSHDEPLPMFTSCCPQWVAMMEKSYPELIPFVSSC	194
Db	132	FGADLTIMESGSELLHRLTEHLEHPSHDEPLPMFTSCCPGWIAMLEKSYDLPPIVSSC	191
Qy	195	KSPQMMGMAMVTKYLSKQGIKAKIIVMVSVMPCVRKQGEADREMFVCV-SEPGVVDVHV	253
Db	192	KSPQMLLAAMVKSYLEAKGIAFKDMVWVSIMPTCKQSEADRMFCVADDTLRQLDQHV	251
Qy	254	ITTAELGNIPKRGINLPDSDWDQPLGLGSGAGVLFGTTCGVMEEALRTAYEIVTKK	313
Db	252	ITTVELGNIPKRGINLAELPEGEWDMVPMGSGAGVLFGTTCGVMEEALRTAYELFTGT	311
Qy	314	PLPRLNLSEVRGDDGKIGASVTIIVPAPGSKFAELVAERLAHKVVEAAAAAEEGAVK	373
Db	312	PLPRLNLSEVRGDDGKIGKTTITVVPAPGSKFEELLKHR-----AAARAEAAAHG-TP	362
Qy	374	PPTAYDGGQGFSTDDGKGLKLVAVANGLNKAKLIGKVVSGEAKYDFVEIMACPGCV	433
Db	363	GPLAWDGGAGFTSDGGRGGITLRVAVANGLNKAKLITKMQAGEAKYDFVEIMACPGCV	422

GenCore version 5.1.6
Copyright (c) 1993 - 2005 CompuGen Ltd.

OM protein - protein search, using sw model

December 21, 2005, 22:55:37 ; Search time 228 Seconds
(without alignments)
1562.683 Million cell updates/sec

US-10-763-712A-26
Pre: 2604
ble: 1 MALGLLAELRAGQAVACARR.....HRAHLLHLYVPGAEADA 505

BLOSUM62
Gapop 10.0 , Gapext 0.5

2166443 seqs, 705528306 residues

per of hits satisfying chosen parameters: 2166443

B seq length: 0

B seq length: 2000000000

essing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

: UniProt_05_80.*

1: uniprot_sprot.*

2: uniprot_trembl.*

red. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
nd is derived by analysis of the total score distribution.

SUMMARIES

Score	Query Match	Length	ID	Description
2604	100.0	505	2 QVZ20	CHLRE
2549	97.9	505	2 QVZ20	CHLRE
2542	97.6	505	2 QVZ20	CHLRE
1702.5	65.4	497	2 QVYU1	CHLRE
1419.5	54.5	436	2 QVX03	CHLRE
1408	54.1	449	2 QVX03	CHLRE
1329	51.0	458	2 QVX03	CHLRE
1225.5	47.1	403	2 QVX03	CHLRE
1016	39.0	608	2 QVX03	CHLRE
973.5	37.4	581	2 QVX03	CHLRE
968	37.2	596	2 QVX03	CHLRE
940	36.1	555	2 QVX03	CHLRE
932.5	35.8	389	2 QVX03	CHLRE
922.5	35.4	636	2 QVX03	CHLRE
902.5	34.7	579	2 QVX03	CHLRE
896.5	34.4	585	2 QVX03	CHLRE
884.5	34.0	588	2 QVX03	CHLRE
874	33.6	456	2 QVX03	CHLRE
862	33.1	471	2 QVX03	CHLRE
859.5	33.0	578	2 QVX03	CHLRE
818.5	31.4	589	2 QVX03	CHLRE
814.5	31.3	606	2 QVX03	CHLRE
814.5	31.3	606	2 QVX03	CHLRE
802	30.8	1198	2 QVX03	CHLRE
791	30.4	582	2 QVX03	CHLRE
789.5	30.3	619	2 QVX03	CHLRE
786	30.2	1206	2 QVX03	CHLRE
784.5	30.1	468	2 QVX03	CHLRE
774	29.7	484	2 QVX03	CHLRE
767	29.5	574	2 QVX03	CHLRE
750.5	28.8	574	1 PHF1	CLOPA

32.	743.5	28.6	596	2	Q73MB6	TREDE	Q73mb6	treponema d
33	741	28.5	572	2	Q9ZNE4	CLOPE	Q9zne4	clostridium
34	739	28.4	572	2	Q9RHU8	CLOPE	Q9rhu8	clostridium
35	716	27.5	582	2	Q59262	CLOAB	Q59262	clostridium
36	702	27.0	562	2	Q5MB2	CLOAB	Q5mb2	clostridium
37	689	26.5	421	2	O8311	DESFR	O8311	desulfovibr
38	660	25.3	421	1	PHFL	DESFR	P07598	desulfovibr
39	639.5	24.6	449	2	Q27096	TRIVA	Q27096	trichomonas
40	636	24.4	421	1	PHFL	DESFR	P13629	desulfovibr
41	635.5	24.4	421	2	Q9AM36	DESDE	Q9am36	desulfovibr
42	624.5	24.0	410	2	O8EAI2	SHEON	O8eai2	shewanella
43	572	22.0	645	2	O52683	THEMA	O52683	thermotoga
44	536	20.6	232	2	Q5Y8D3	9CLOT	Q5y8d3	clostridium
45	517	19.9	232	2	Q5Y8D2	9CLOT	Q5y8d2	clostridium

ALIGNMENTS

RESULT 1

QVZ20 CHLRE PRELIMINARY; PRT; 505 AA.

AC QVZ20

DT 01-MAR-2002 (Tremblrel. 20, Created)

DT 01-MAR-2002 (Tremblrel. 20, Last sequence update)

DT 10-MAY-2005 (Tremblrel. 30, Last annotation update)

DE Iron-hydrogenase HydA2 (Fe-only hydrogenase precursor).

GN Name=hydA2; Synonyms=HYDA2;

OS Chlamydomonas reinhardtii.

OC Eukaryota; Viridiplantae; Chlorophyta; Chlorophyceae;

OC Chlamydomonadales; Chlamydomonadaceae; Chlamydomonas.

OX NCBI_TaxID=3055;

RN [1]

RP NUCLEOTIDE SEQUENCE

RC STRAIN=21gr, and Cc425;

RX PubMed=12823545; DOI=10.1046/j.1432-1033.2003.03656;

RA Forestier M., King P., Zhang L., Posewitz M., Schwarzer S., Happe T.,

RA Ghirardi M.L., Seibert M.

RT "Expression of two [Fe]-hydrogenases in Chlamydomonas reinhardtii

RT under anaerobic conditions."

RL Eur. J. Biochem. 270:2750-2758(2003).

DR EMBL; AY055756; AAU23573.1; -; mRNA.

DR EMBL; AY090770; ANM01186.2; -; Genomic_DNA.

DR HSSP; P29166; 1FEH.

DR GO; GO:0005489; F:electron transporter activity; IEA.

DR GO; GO:0008901; F:ferridoxin hydrogenase activity; IEA.

DR GO; GO:0006118; F:electron transport; IEA.

DR InterPro; IPR004108; Fe_hyd_1g_C.

DR InterPro; IPR003149; Fe_hyd_SSU.

DR Pfam; PF02256; Fe_hyd_SSU; 1.

KW Transit peptide.

FT TRANSIT

SQ SEQUENCE 505 AA; 53721 MW; 3A1948749B033EA9 CRC64;

Query Match 100.0%; Score 2604; DB 2; Length 505;

Best Local Similarity 100.0%; Pred. No. 7.1e-166; Indels 0; Gaps 0;

Matches 505; Conservative 0; Mismatches 0;

Qy 1 MALGLLAELRAGQAVACARRTNAPAHVAVVPCLPSPRAGKFFNLSQKVPSSQSGSTIR 60

Db 1 MALGLLAELRAGQAVACARRTNAPAHVAVVPCLPSPRAGKFFNLSQKVPSSQSGSTIR 60

Qy 61 VAATATDAVPHWKALBELDKPKDGGKVLIAQVAVPAVVAIAESFGLAPAVSPKLAT 120

Db 61 VAATATDAVPHWKALBELDKPKDGGKVLIAQVAVPAVVAIAESFGLAPAVSPKLAT 120

Qy 121 GLRALGPDQVFDTLFAADLTIMBEETLLHRLKEHLEAHPHSDPLPMFTSCCPGWAVM 180

Db 121 GLRALGPDQVFDTLFAADLTIMBEETLLHRLKEHLEAHPHSDPLPMFTSCCPGWAVM 180

Qy 181 EKSPYELIPFVSSCKSPQMMGMAMVKTLYLSEKQICPAKDIWVSVMPVCRVQSGADREWF 240

GenCore version 5.1.6
Copyright (c) 1993 - 2005 Compugen Ltd.

OM protein - protein search, using sw model

December 21, 2005, 22:55:37 : Search time 228 seconds

	32	743.5	28.6	596	2	Q73MB6_TREDE
	33	741	28.5	572	2	Q9ZNE4_CLOPE
	34	739	28.4	572	2	Q9RHU8_CLOPE
	35	716	27.5	582	2	Q59262_CLOAB
	36	702	27.0	562	2	Q5MIB2_9CLOT
	37	689	26.5	421	2	O08311_DESFR
	38	660	25.3	421	1	PHF1_DESVH
	39	639.5	24.6	449	2	Q27096_TRIVA
						SHPT_DRSVO

Q73mb6 treponema d
Q9zne4 clostridium
Q9rhu8 clostridium
Q59262 clostridium
Q5mib2 clostridium
O08311 desulfovibr
P07598 desulfovibr
Q27096 trichomonas
P13629 desulfovibr
Q9am36 desulfovibr
Q8eai2 shewanella
O52683 thermotoga
Q5y8d3 clostridium
Q5y8d2 clostridium

Protein Sequence Searches - February 2005

All of the sequence databases on ABSS have recently been updated.

- Please note that the curators of the UniProt database have purged some temporary accession numbers from the most recent version of UniProt. These sequences have been assigned new permanent accession numbers. The new UniProt record may not contain the previous temporary accession number.
- If you encounter an accession number from an older search run against UniProt (results file extension .rup) that can no longer be found in the database, the permanent record with the new accession number can be found by searching the old accession number in the UniProt-Protein Archive database (UniPARC) at:

<http://www.pir.uniprot.org/database/archive.shtml>

If you have any questions regarding this information or your results, please contact any SMC searcher.

When submitting sequence search results for scanning into IFW, please include a copy of this attachment to assist any future Examiners or members of the public who may encounter UniProt temporary accession numbers.

27	786	30.2	1206	2	Q27094_TRIVA
28	784.5	30.1	468	2	Q27094_TRIVA
29	774	29.7	484	2	Q9RGN3_MEGEL
30	767	29.5	574	2	Q59261_CLOSA
31	750.5	28.8	574	1	PHF1_CLOPA

Q27094 trichomonas
Q9rgn3 megasphaera
Q59261 clostridium
P29166 clostridium

QY

181 EKSYBELIPFVSCKSPQMMGANVTYLSKOGIPAKDIIWVSVMPCVRRKQGRADREWF

|||||

AA.

update)
in update)
ie precursor).

.orophyceae;
.amydomonas.

2003.03656;
M., Schwarzer S., Happe T.,
alamydomonas reinhardtii

A.

tivity; IEA.
activity; IEA.

033EA9 CRC64;

DB 2; Length 505;
.1e-166;
8 0; Indels 0; Gaps 0;

CLPSRAGKFFNLQKVPSSQSARGSTIR 60

CLPSRAGKFFNLQKVPSSQSARGSTIR 60

QVAPAVRVAIAESFGLAPGVSPGKLAT 120

QVAPAVRVAIAESFGLAPGVSPGKLAT 120

KEHLEAHPHSDPEPLPMFTSCCFGWVAMM 180

KEHLEAHPHSDPEPLPMFTSCCFGWVAMM 180

GenCore version 5.1.6
Copyright (c) 1993 - 2005 Compugen Ltd.

OM protein - protein search, using sw model

Run on: December 21, 2005, 22:56:32 ; Search time 39 Seconds

(without alignments)
1245.884 Million cell updates/sec

Title: US-10-763-712A-26

Perfect score: 2604

Sequence: 1 MALGGLAELRAGQAVACARR.....HRAHELLTHYVPGAEADA 505

Scoring table:

BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 283416 seqs, 96216763 residues

Total number of hits satisfying chosen parameters: 283416

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

PIR 80:*

1: pir1:*

2: pir2:*

3: pir3:*

4: pir4:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1016	39.0	608	2 C72405	hydrogenase (EC 1.1.1.1)
2	896.5	34.4	585	2 D57150	hydrogenase (EC 1.1.1.1)
3	814.5	31.3	606	2 S13526	hydrogenase (EC 1.1.1.1)
4	786	30.2	1206	2 T18557	probable hydrogenase
5	750.5	28.8	574	1 HQCLLP	hydrogenase (EC 1.1.1.1)
6	716	27.5	582	2 D96903	hydrogenase (EC 1.1.1.1)
7	716	27.5	582	2 JC6002	hydrogenase (EC 1.1.1.1)
8	660	25.3	421	1 HQDVFL	cytochrome-c3 hydr
9	636	24.4	421	1 HQDVFL	cytochrome-c3 hydr
10	572	22.0	645	2 G72256	hydrogenase (EC 1.1.1.1)
11	339.5	13.0	538	2 T40992	hypothetical prote
12	274.5	10.5	491	2 S63206	LET1 protein homol
13	260	10.0	469	2 S31336	LET1 protein - yea
14	226	8.7	450	2 B97297	hydrogenase chain
15	127	4.9	124	1 HQDVSV	hydrogenase (EC 1.1.1.1)
16	126.5	4.9	301	2 B72256	hypothetical prote
17	118	4.5	916	2 F97053	penicillin-binding
18	117.5	4.5	123	1 HQDVFS	hydrogenase (EC 1.1.1.1)
19	115.5	4.4	903	1 C64444	cell division cont
20	114	4.4	574	1 KIHUPL	pyruvate kinase (E
21	114	4.4	838	2 B3150	probable ATP-depen
22	114	4.4	4077	2 T17484	hypothetical prote
23	113	4.3	732	2 A69086	cell division cont
24	111.5	4.3	1118	2 T27865	hypothetical prote
25	111.5	4.3	2117	2 T36180	CDA peptide synthe
26	110.5	4.2	733	2 T35429	probable fatty oxi
27	109	4.2	503	2 F95988	probable ATP-depen
28	108	4.1	344	2 H64477	malate dehydrogena
29	108	4.1	468	2 E83046	L-seryl-tRNA ^{Sec} se

30 108 4.1 4151 2 G70944 probable polyketid
31 107 4.1 943 2 T34847 probable transcrip
32 106.5 4.1 426 2 B84261 UDP-glucose dehydr
33 106.5 4.1 478 2 E75359 glycolate oxidase
34 106.5 4.1 543 1 KIHUPL pyruvate kinase (E
35 106.5 4.1 2126 2 H70621 probable polyketid
36 106 4.1 512 2 A98352 hydantoinase homol
37 106 4.1 512 2 A98352 hydantoinase A [im
38 106 4.1 757 2 S68142 probable transcrip
39 106 4.1 823 2 T35280 probable integral
40 106 4.1 1461 2 B70588 probable polyketid
41 105.5 4.1 631 2 B87250 dnaK protein (impo
42 105.5 4.1 907 2 AD2951 cell division prot
43 105.5 4.1 910 2 H98331 cell division prot
44 105 4.0 815 2 T35970 probable efflux pr
45 105 4.0 3739 2 T17410 polyketide synthas

ALIGNMENTS

RESULT 1

C72405

hydrogenase (EC 1.18.99.1) (Fe) large chain [similarity] - Thermotoga maritima (strain
C:Species: Thermotoga maritima
C:Date: 11-Jun-1999 #sequence_revision 11-Jun-1999 #text_change 16-Aug-2004
C:Accession: C72405
R:Nelson, K.E.; Clayton, R.A.; Gill, S.R.; Gwinn, M.L.; Dodson, R.J.; Haft, D.H.; Hick
Garrett, M.M.; Stewart, A.M.; Cotton, M.D.; Pratt, M.S.; Phillips, C.A.; Richardson, D.
C.M.

Nature 399, 323-329, 1999

A:Title: Evidence for lateral gene transfer between Archaea and Bacteria from genome s

A:Reference number: A72200; MUID:99287316; PMID:10360571

A:Accession: C72405

A:Status: preliminary

A:Molecule type: DNA

A:Residues: 1-608 <ARN>

A:Cross-references: UNIPROT:Q9WY44; UNIPARC:UPI0000003A72; GB:AE001705; GB:AE000512; N

A:Experimental source: strain MSB8

C:Genetics:

A:Gene: TM0201

C:Superfamily: Hydrogenase (Fe) large chain; ferredoxin 2[4Fe-4S] homology

C:Keywords: 2Fe-2S; 4Fe-4S; iron; metalloprotein; oxidoreductase

F:162,74,77,91/Binding site: 2Fe-2S cluster (Cys) (covalent) #status predicted

F:123,127,130,136/Binding site: 4Fe-4S cluster (His, Cys, Cys) (covalent) #status predicted

F:174,177,180,227/Binding site: 4Fe-4S cluster (Cys) (covalent) #status predicted

F:184,217,220,223/Binding site: 4Fe-4S cluster (Cys) (covalent) #status predicted

F:329,384,528,532/Binding site: 4Fe-4S cluster (Cys) (covalent) #status predicted

F:532/Binding site: diiron cofactor (Cys) #status predicted

Query Match 39.0%; Score 1016; DB 2; Length 608;

Best Local Similarity 48.7%; Pred. No. 1.5e-63;

Matches 207; Conservative 59; Mismatches 91; Indels 68; Gaps 6;

QY 73 KLAELDKPKDGRKVLIAQVAPAVRVAIAESFGLAGAVSPGKLAGLALGDFQVDF 132

DB 239 KVLLELEKKE- - - - - KILVVQTAPSVRVAIGSEFGYAPGTISTGQVAAALRELGLDFYDF 294

QY 133 TLFAADLTMEEGTELLHRLKEHLEAHPHSD-EPUPMTFTSCCPGVVAMMEKSYPELIPFV 191

DB 295 TNFGADLTMEEGSEFLEK- - - - - GDUEDLPMTFTSCCPGVVNLVKRYPELIRL 347

QY 192 SSCKSPQMMGMVKTYLSEKQIPAKDIVMVSVMPCVRKQGEADREWFVCSPEGVDFD 251

DB 348 SSACKSPQMLSAMVKTYFAELKGVKPEDFHVSIMPTAKKDEALRKLQMLVN- -GVPAVD 405

QY 252 HVITTAELGNIFKRGINLPELPSDWDQPLGLSGAGVLFGTTGGVMEALRTAYEIVT 311

DB 406 VVLTATRELGLIRMKKIPFANLPBEEVDAPLGISTGAAALFGVTGGVMEALRTAYELK 465

QY 312 KEPLPRMLSEVRGLDGIKEASVTLVPAPGSKFAELVAERLAHKVVEAAAAAEEGA 371

DB 466 GKALPKIVFEVRGLKGVREABIDL- - - - - 490

GenCore version 5.1.6
Copyright (c) 1993 - 2005 Compugen Ltd.

OM protein - protein search, using sw model

Run on: December 21, 2005, 22:54:57 ; Search time 186 Seconds

(without alignments)
1192.937 Million cell updates/sec

Title: US-10-763-712A-26

Perfect score: 2604

Sequence: 1 MALGLLAEIRAGQAVACARR.....HRAELLTHVPGAEADA 505

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 2443163 seqs, 439378781 residues

Total number of hits satisfying chosen parameters: 2443163

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : A_Geneseq_21.*

- 1: Geneseqp1980s.*
- 2: Geneseqp1990s.*
- 3: Geneseqp2000s.*
- 4: Geneseqp2001s.*
- 5: Geneseqp2002s.*
- 6: Geneseqp2003as.*
- 7: Geneseqp2003bs.*
- 8: Geneseqp2004s.*
- 9: Geneseqp2005s.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	2604	100.0	505	9 AEB73240	Aeb73240 Segment o
2	2598	99.8	505	9 AEB73337	Aeb73337 Gas Chann
3	2549	97.9	505	9 AEB73302	Aeb73302 Segment o
4	2542	97.6	505	9 AEB73303	Aeb73303 Segment o
5	2005	77.0	393	8 ADU00340	ADU00340 Iron hydr
6	1702.5	65.4	497	9 ADY27749	ADY27749 Chlamydom
7	1702.5	65.4	497	9 AEB73305	Aeb73305 Segment o
8	1702.5	65.4	497	9 AEB73328	Aeb73328 Segment o
9	1483.5	57.0	386	8 ADU00341	ADU00341 Iron hydr
10	1458.5	56.0	388	8 ADU00342	ADU00342 Iron hydr
11	1419.5	54.5	436	9 ADY27750	ADY27750 Chlorella
12	1419.5	54.5	436	9 AEB73219	Aeb73219 Segment o
13	1412	54.2	415	9 AEB73329	Aeb73329 Segment o
14	1408	54.1	449	9 AEB73297	Aeb73297 Segment o
15	1400.5	53.8	448	9 ADY27748	ADY27748 Scenedes
16	1225.5	47.1	403	9 AEB73241	Aeb73241 Segment o
17	1016	39.0	608	9 AEB73230	Aeb73230 Segment o
18	1016	39.0	608	9 AEB73304	Aeb73304 Segment o
19	973.5	37.4	581	9 AEB73306	Aeb73306 Segment o
20	973.5	37.4	581	9 AEB73242	Aeb73242 Segment o
21	940	36.1	555	9 AEB73244	Aeb73244 Segment o
22	932.5	35.8	389	9 AEB73245	Aeb73245 Segment o
23	922.5	35.4	636	9 AEB73243	Aeb73243 Segment o
24	922.5	35.4	636	9 AEB73307	Aeb73307 Segment o

25	909.5	34.9	506	9 AEB73300	Aeb73300 Segment o
26	907	34.8	580	6 AEB73308	Aeb73308 Segment o
27	902.5	34.7	579	9 AEB73308	Aeb73308 Segment o
28	896.5	34.4	585	9 AEB73228	Aeb73228 Segment o
29	896.5	34.4	585	9 AEB73310	Aeb73310 Segment o
30	884.5	34.0	588	9 AEB73309	Aeb73309 Segment o
31	881	33.8	554	9 AEB73247	Aeb73247 Segment o
32	873	33.5	458	9 AEB73246	Aeb73246 Segment o
33	859.5	33.0	578	9 AEB73248	Aeb73248 Segment o
34	818.5	31.4	589	9 AEB73312	Aeb73312 Segment o
35	818.5	31.4	589	9 AEB73236	Aeb73236 Segment o
36	814.5	31.3	606	9 AEB73223	Aeb73223 Segment o
37	814.5	31.3	606	9 AEB73311	Aeb73311 Segment o
38	791	30.4	582	7 ADC59546	Adc59546 Clostridi
39	791	30.4	582	8 ADR90318	Adr90318 Clostridi
40	789.5	30.3	619	9 AEB73249	Aeb73249 Segment o
41	786	30.2	1206	9 AEB73232	Aeb73232 Segment o
42	786	30.2	1206	9 AEB73313	Aeb73313 Segment o
43	784.5	30.1	468	9 AEB73234	Aeb73234 Segment o
44	784.5	30.1	468	9 AEB73314	Aeb73314 Segment o
45	774	29.7	484	9 AEB73226	Aeb73226 Segment o

ALIGNMENTS

RESULT 1
AEB73240
ID AEB73240 standard; protein; 505 AA.
XX
AC AEB73240;
XX
DT 20-OCT-2005 (first entry)
XX
DE Segment of microbial iron hydrogenase, SEQ ID NO 26.
XX
KW Genetic engineering; biogas; fuel; screening;
KW Genetically engineered microorganism; site-directed mutagenesis;
KW Cell culture; iron hydrogenase; enzyme.
XX
OS Chlamydomonas reinhardtii.
XX
PN WO2005072262-A2.
XX
PD 11-AUG-2005.
XX
PF 21-JAN-2005; 2005WO-US001983.
XX
PR 21-JAN-2004; 2004US-00763712.
XX
(SOLA-) SOLA2YME INC.
PI Dillon HF;
XX
WPI; 2005-564413/57.
XX
Engineering a cell to produce an increased amount of hydrogen by transforming a cell with the mutagenized nucleic acid sequence and screening or selecting the cell for an increased amount of hydrogen.
XX
Claim 11; SEQ ID NO 26; 297pp; English.
XX
The present invention relates to a method for engineering a cell to produce an increased amount of hydrogen by providing a mutagenized nucleic acid derived from a gene that encodes a protein involved in hydrogen production; transforming a cell with the mutagenized sequence; and screening or selecting the cell for an increased amount of hydrogen. Also claimed is a method of multi-parental mating of microbes that mate in response to a stimulus. The first gene encodes ferredoxin, catalase, isocitrate lyase, malate dehydrogenase, 14-3-3 protein, enolase, aldolase, ribosomal protein S8, ribosomal protein L17, ribosomal protein S18, ribosomal protein L37, ribosomal protein L12, ribosomal protein S15, iron hydrogenase, nickel-iron hydrogenase or components of the photosystem I,

GenCore version 5.1.6
Copyright (c) 1993 - 2005 Compugen Ltd.

QM protein - nucleic search, using frame_plus_p2n model
Run on: December 25, 2005, 02:18:02 ; Search time 175 Seconds
(without alignments)
1497.555 Million cell updates/sec

Title: US-10-763-712A-26
Perfect score: 2604
Sequence: 1 MALLGLAELRAGAVACARR.....HRAHELLHTYVPGAEADA 505

Scoring table: BLOSUM62
Xgapop 10.0, Xgapext 0.5
Ygapop 10.0, Ygapext 0.5
Fgapop 6.0, Fgapext 7.0
Delop 6.0, Delext 7.0

Searched: 4168288 seqs, 259477437 residues

Total number of hits satisfying chosen parameters: 8336576

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Command line parameters:
-MODEL=frame+ p2n.model -DEV=xlh
-Q=/cgn2_1/USPTO.spool/US10763712/runat 21122005 164206 10330/app query.fasta 1.647
-DB=Published Applications NA New -QFMT=fastap -SUFFIX=p2n.rnpbn -MINMATCH=0.1
-LOOPL=0 -LOOPEXT=0 -UNITS=bits -START=1 -END=1 -MATRIX=blosum62
-TRANS=human40.cdi -LIST=45 -DOCALIGN=200 -THR SCORE=pct -THR MAX=100
-THR MIN=0 -ALIGN=15 -MODE=LOCAL -OUTFMT=ptc -HEAPSIZ=500 -MINLEN=0
-MAXLEN=200000000 -USER=US10763712 @CGN 1 1 122 @runat 21122005 164206 10330
-NCPU=6 -ICPU=3 -NO MAP -LARGQUERY -NEG SCORES=0 -WAIT -DSPBLOCK=100
-LONGLOG -DEV TIMEOUT=120 -WARN TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5
-FGAPOF=6 -FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database : Published Applications NA New:

- 1: /cgn2_6/ptodata/2/pubpna/US08_NEW_PUB.seq:
- 2: /cgn2_6/ptodata/2/pubpna/US06_NEW_PUB.seq:
- 3: /cgn2_6/ptodata/2/pubpna/US07_NEW_PUB.seq:
- 4: /cgn2_6/ptodata/2/pubpna/PCT_NEW_PUB.seq:
- 5: /cgn2_6/ptodata/2/pubpna/US09_NEW_PUB.seq:
- 6: /cgn2_6/ptodata/2/pubpna/US10_NEW_PUB.seq:
- 7: /cgn2_6/ptodata/2/pubpna/US11_NEW_PUB.seq:
- 8: /cgn2_6/ptodata/2/pubpna/US11_NEW_PUB.seq:
- 9: /cgn2_6/ptodata/2/pubpna/US11_NEW_PUB.seq:
- 10: /cgn2_6/ptodata/2/pubpna/US60_NEW_PUB.seq:

pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	1707	65.6	1494	US-10-763-712A-173	Sequence 173, App
2	1419.5	54.5	1311	US-10-763-712A-178	Sequence 178, App
3	1408	54.1	1350	US-10-763-712A-177	Sequence 177, App
4	750.5	28.8	1725	US-10-763-712A-174	Sequence 174, App
5	672.5	25.8	1265	US-10-763-712A-175	Sequence 175, App
6	472	18.1	1407	US-10-763-712A-176	Sequence 176, App
7	132	5.1	4549	US-10-821-234-316	Sequence 316, App
8	124	4.8	3079	US-10-623-155-116	Sequence 116, App

9	113.5	4.4	168516	7	US-11-121-086-3	Sequence 3, Appli
10	112	4.3	3408	6	US-10-858-730-40	Sequence 40, Appl
11	108.5	4.2	11438	6	US-10-821-234-224	Sequence 224, App
12	108.5	4.2	78869	7	US-11-075-185-1	Sequence 1, Appli
13	107	4.1	1545	7	US-11-055-822-431	Sequence 431, App
14	107	4.1	9066	7	US-11-129-143-42	Sequence 42, Appl
15	107	4.1	9066	7	US-11-129-143-44	Sequence 44, Appl
16	107	4.1	9066	7	US-11-129-143-46	Sequence 46, Appl
17	107	4.1	9066	7	US-11-129-143-48	Sequence 48, Appl
18	107	4.1	9066	7	US-11-129-143-50	Sequence 50, Appl
19	107	4.1	9066	7	US-11-121-086-29	Sequence 29, Appl
20	107	4.1	160226	7	US-11-121-086-29	Sequence 29, Appl
21	106	4.1	10968	7	US-11-075-185-35	Sequence 35, Appl
22	106	4.1	116856	7	US-11-143-980-1	Sequence 1, Appli
23	104.5	4.0	841	6	US-10-750-185-31039	Sequence 31039, A
24	104.5	4.0	2697	7	US-11-174-150-19	Sequence 19, Appl
25	103	4.0	164810	7	US-11-121-086-4	Sequence 4, Appli
26	103	4.0	172543	7	US-11-121-086-6	Sequence 6, Appli
27	102.5	3.9	150468	7	US-11-112-908-56	Sequence 56, Appl
28	102.5	3.9	133789	7	US-11-112-908-55	Sequence 55, Appl
29	102	3.9	1545	6	US-10-467-657-1967	Sequence 1967, Ap
30	101	3.9	1338	6	US-10-067-374-11	Sequence 11, Appl
31	101	3.9	1400	7	US-11-055-822-45	Sequence 45, Appl
32	100.5	3.9	16933	6	US-10-995-561-13257	Sequence 13257, A
33	100.5	3.9	31320	6	US-10-995-561-13309	Sequence 13309, A
34	100.5	3.9	98345	7	US-11-112-908-36	Sequence 36, Appl
35	100.5	3.9	127340	7	US-11-112-908-35	Sequence 35, Appl
36	100	3.8	2262	7	US-11-137-465-18	Sequence 18, Appl
37	99.5	3.8	78869	7	US-11-075-185-1	Sequence 1, Appli
38	99	3.8	10524	7	US-11-075-185-37	Sequence 37, Appl
39	98.5	3.8	2829	6	US-10-750-185-26265	Sequence 26265, A
40	98.5	3.8	26277	6	US-10-995-561-13414	Sequence 13414, A
41	98	3.8	1080000	6	US-10-928-446A-1	Sequence 1, Appli
42	98	3.8	1080000	6	US-10-928-446A-181	Sequence 181, App
43	98	3.8	1080000	6	US-10-928-446A-183	Sequence 183, App
44	98	3.8	1080000	6	US-10-928-446A-185	Sequence 185, App
45	98	3.8	1080000	6	US-10-928-446A-187	Sequence 187, App

ALIGNMENTS

RESULT 1

US-10-763-712A-173
; Sequence 173, Application US/10763712A
; Publication No. US20050266541A1
; GENERAL INFORMATION:
; APPLICANT: Solazyme, Inc.
; APPLICANT: Dillion, Harrison F.
; TITLE OF INVENTION: Methods and Compositions for Evolving Microbial Hydrogen
; TITLE OF INVENTION: Production
; FILE REFERENCE: H2042101-CIP
; CURRENT APPLICATION NUMBER: US/10763,712A
; CURRENT FILING DATE: 2004-01-21
; PRIOR APPLICATION NUMBER: US 10/287,750
; PRIOR FILING DATE: 2002-11-04
; PRIOR APPLICATION NUMBER: US 10/411,910
; PRIOR FILING DATE: 2003-04-12
; PRIOR APPLICATION NUMBER: US 60/500,032
; PRIOR FILING DATE: 2003-09-03
; NUMBER OF SEQ ID NOS: 184
; SOFTWARE: Patentin version 3.2
; SEQ ID NO 173
; LENGTH: 1494
; TYPE: DNA
; ORGANISM: Chlamydomonas reinhardtii

Alignment Scores:
Pred. No.: 3.09e-150 Length: 1494
Score: 1707.00 Matches: 338
Percent Similarity: 77.06% Conservatives: 55
Best Local Similarity: 66.27% Mismatches: 75
Query Match: 65.55% Indels: 42

**This Page is Inserted by IFW Indexing and Scanning
Operations and is not part of the Official Record**

BEST AVAILABLE IMAGES

Defective images within this document are accurate representations of the original documents submitted by the applicant.

Defects in the images include but are not limited to the items checked:

- ☐ BLACK BORDERS
- ☒ IMAGE CUT OFF AT TOP, BOTTOM OR SIDES
- ☐ FADED TEXT OR DRAWING
- ☒ BLURRED OR ILLEGIBLE TEXT OR DRAWING
- ☐ SKEWED/SLANTED IMAGES
- ☐ COLOR OR BLACK AND WHITE PHOTOGRAPHS
- ☐ GRAY SCALE DOCUMENTS
- ☐ LINES OR MARKS ON ORIGINAL DOCUMENT
- ☐ REFERENCE(S) OR EXHIBIT(S) SUBMITTED ARE POOR QUALITY
- ☐ OTHER: _____

IMAGES ARE BEST AVAILABLE COPY.

As rescanning these documents will not correct the image problems checked, please do not report these problems to the IFW Image Problem Mailbox.